State of Michigan

Department of Information Technology

Achieving Breakthrough Data Transformation through a Shared Solution

Category: “Data, Information and Knowledge Management”

B. Executive Summary:

Managing data and information has become a growing concern for many Information Technology (IT) units. As most government agencies are acutely aware, the public sector is rich with data. Governmental units are known for collecting and storing vast quantities of spatial, financial, health and human services, address, judicial and corrections, tax, education and management data.

Maintaining the quality and usefulness of this data often falls on aging systems and a workforce where many are facing retirement in the near future. Many of the systems used to load and transfer these vast data stores have been around for decades and typically lack sufficient documentation to allow newer workers to understand the data handling. Further, file structures often contain large amounts of redundant data records that must be reviewed constantly to minimize unnecessary duplication.

The State of Michigan, known for its innovation in the area of IT and in the use of shared services to run its Center for Geographic Information Systems, is leveraging that model into a new Center for Shared Solutions (CSS). The CSS approached a major Extract-Transfer-Load (ETL) tool vendor to negotiate converting a limited use license into an enterprise license, which could be redeployed to address the data management issues facing the state—**improving data quality, saving costs, speeding delivery and updating aging technologies.**

Data quality improvements, based on responses from an implementation within the Department of Treasury range from 10-25% over existing processes. These improvements will show themselves in improved matching of tax non-compliance activities and recovery efforts. This will result in 500-1,500 fewer incorrect tax audits and a potential additional $13-$33 Million in tax revenue.

Now having a standard and centralized information management service allows the state to better manage its data and information stores. With built-in data quality and data profiling services, the state can gather information about its data much faster than conventional methods. This helps even the newest developer better understand the data that they are charged to manage.

In addition, because the CSS model makes a well-trained and highly-qualified staff of resources available across state agencies, business units are able to begin work on data quality improvement projects much quicker than contracting out or training internal resources for the work. Using the CSS business model and the information management tools provides huge benefits in terms of data quality, cost savings, delivery speed, and tool modernization. This translates into better, more efficient government in Michigan.

C. Description:

Problem: Since the 1960's, Michigan state government has collected, managed and distributed information in thousands of databases, data marts and warehouses. During this time, these data stores have been used to monitor public health, manage revenues/taxes, manage state resources, provide state assistance, protect the health and welfare of children, manage infrastructure assets, and provide for public safety and education. The information management *within* these systems has been very successful as long as their operational boundaries did not change much, but change has become an ever-increasing challenge. The need for more information to make better policy decisions amidst a constantly changing legal, administrative, business and economic climate has forced information integration to become commonplace. Moving, validating and integrating information among systems is often an under-planned, but critical step in the IT integration process. While the process of information management between systems is a critical success factor, it has been one of the least managed processes in the state.

Over the course of several decades, thousands of custom-coded programs were developed in the state to move data between systems in a "home-grown" extract-transfer-load (ETL) data management environment. These custom coded ETL programs often lacked up-to-date supporting documentation and, over time, were patched and modified numerous times. The resulting ETL processes often required hours of maintenance and oversight in order to complete run cycles. The data being processed through this code often contained invalid values and missing or incorrect information. This substantially reduced the **quality of data necessary to efficiently and effectively conduct business**. As a result, one department within the Department of Treasury had fallen several years behind in the loading of IRS data used for matching and auditing tax non-compliance. In addition to reporting, high numbers of job failures have resulted from poor quality data.

During 2008, the Michigan Department of Information Technology (MDIT) noted that several large, new projects were employing ETL services from various vendors to assist with the data conversion/integration of major systems. Unfortunately, these applications and services were purchased with limited licenses and could only be used for the purposes identified in the contracts. In addition, a full license and the associated hardware to run such software is cost-prohibitive for any one agency to pursue.

Solution: In March of 2008, the Center for Geographic Information (CGI) at the State of Michigan leveraged its business model into the shared service arena and created a Center for Shared Solutions (CSS). This model views the department as a cost center that sells its services to other State agencies to fund its business. The department then implemented a rated-service model for the applications and services. Finally, the department is resourced to provide staff augmentation services to assist departments with various projects.

Starting in July of 2008, the newly created CSS negotiated with IBM to turn a limited use Extract-Transfer-Load (ETL) license into a high-powered enterprise solution to manage

the information integration process. Several enterprise grade servers were purchased for the software, capitalizing the expenses over three years. The goals of this new team were to establish a shared service for use within the MDIT that could be leveraged across all agencies to **improve the quality and performance of information management through the ETL process.** In the nine months following the production of the service, significant benefits have been realized.

D. Significance:

***Improved operation**s* - This project is pivotal to implementing the second goal of the MDIT Strategic Plan which is to "Create efficiencies for our agencies in support of their existing systems." These efficiencies include the **improvement in the quality and reliability of the data** that is being processed. The creation of the Shared Solutions Team and the implementation of ETL service area is a keystone for this aspect of the MDIT strategic plan.

This project also emphasizes the Governor's Better Government Initiative: *Increased use of innovative technology* – measured by the increase in the number of cross-agency technology initiatives; *increased number of shared services* – measured by the number of transactions processed centrally and the number of departments sharing staff for common functions and; *reduced expenditures for contractual services* – including reduced duplication of maintenance contracts, software and hardware purchases, and on-going maintenance costs. All of these serve to reduce the cost of government to the taxpayer while modernizing the tools used by MDIT and making agency processing more efficient.

Innovative approach - CSS creatively negotiated with a vendor to turn a limited use license into a powerful enterprise information knowledge management solution that would be out of the financial reach of individual agencies or departments. Further, the use of a core team of skilled knowledge workers frees up having to train and maintain specialized skills in the various agency support teams. This has proven to be highly cost-effective while providing a high level of service to the participating agencies as well as creating a solution to address the ongoing dilemma of aging legacy systems.

NASCIO alignment - The staffing and governance model being used by this project is one that has a great potential to be readily transferred to other states agencies. Specifically, the project aligns with the NASCIO priority list for 2009 through consolidation, shared services, budget and cost control, security, ERP strategy, and governance. *Consolidation* - by having the ETL toolset adopted as the standard tool across the state. Efforts are underway to actively market the tool to other agencies within the state, leveraging existing customers' testimonies as reason for adoption. *Shared services*–through use of the existing CGI business model to allow for the centralized purchase and installation of the service and staffing of the model to allow for staff augmentation and supplementation. *Budget and Cost Control* - by giving departments the freedom to utilize the service to maximize the existing staff without having to assume overhead of installing and maintaining the service individually and they do not need to over-staff teams to handle peaks in demand. *Security* - by using the

CSS to allow each department to recognize audit requirements for separation of duties involving production implementation of ETL code. Further, service does not require any data to be stored on the application server, which provides additional security and controls. Audit reports from the tool are standard and additional reports are available providing improved tracking and reporting over existing legacy code. *ERP Strategy* - is provided by the CSS ETL tool in its scalable design in hardware and software and the use of the CSS team as a staff augmentation resource to allow departments to run leaner while maintaining a trained pool of ready resources. Finally, *Governance* - by using the CSS to act as a single point of contact for the ETL toolset. Decision making and coordination is improved and the ability to implement other data governance processes (such as data sharing agreements and common citizen files) is increased.

Impacts - The ability for development teams to leverage a common, centralized service administration team enables a focus on development efforts while not requiring another DBA-level staff member to support the toolset. Development time has been significantly reduced for work being done with the new toolset. Since the product has the capability of generating final documentation, the documentation of the system is always current, replacing the non up-to-date documentation from the legacy system. This self-documenting process also allows a quick analysis of new data elements, modified data elements and other data management activities to be done within the boundaries of the tool. Agency and IT risks are reduced in a number of areas, including accurate documentation, improved testing, and improved data quality; and undocumented custom code is being retired. For Treasury, the ability to deliver the revenue estimates more timely and accurately has improved budget decision-making by the legislature and governor's office. Also, Treasury has used the improved data matching processes to correctly identified delinquent taxpayers in the state.

Scope - Currently this tool is being used by several business units within both the Michigan Department of Treasury and Michigan Department of Community Health. Other state agencies that are currently adopting the data and information management tool include Transportation, Energy, Labor and Economic Growth, Human Services, and Corrections. This project has encompassed several hundred database tables involving billions of records, within its first eight months. Completed agency projects vary from simple database loads to intensive data cleansing and manipulation processes.

E. Benefits of the Project:

Impact and Outcomes - Following the agreement to purchase the software, the CSS began looking for opportunities to evangelize departments that were struggling with custom coded ETL processes and helping them to see the benefits of using the enterprise tool. The Treasury department asked to see what the tool capabilities were in late September 2008. Over the course of two weeks the CSS team, with the help of IBM, was able to replicate the output from an existing application using the ETL tool, complete with testing. The first job from the Information Returns Master File Load process, which normally took about six hours to complete (if it ran without aborting) ran in three minutes using the new software. A second process improved from several hours to minutes. This demonstration was convincing enough for the Department of

Treasury to engage the CSS team in the next challenge. By early October the tool was being utilized to profile data conversions and assist with a major conversion effort, identifying data elements that had values and formats that were incorrect. **An additional 10-25% improvement in data quality, after having been "scrubbed" as part of the conversion effort, was recognized as a result of the data profiling**. This saved the state testing time and helped a major Treasury project meet target dates. In December 2009, the Tax compliance section within Treasury began using the new tool for the development of several other major projects. In order to meet a looming deadline, the Treasury group asked the CSS team to exercise another feature unique to the CSS model, staff augmentation services. Using the ETL tool and a combination of state and contract personnel, the CSS team took the Michigan Business Tax project, estimated at 1500 hours to develop, and completed the development work that included file loads for 51 separate tables in less than 300 hours. This saved the project, Treasury, and the taxpayers of Michigan, $200,000 in cost and several months in effort. Further, the CSS/ETL team was able to **meet a critical revenue estimation date, providing much more accurate and timely information than had been available in the past**. In addition, other projects including processes to data cleanse and load Military Combat Zone files, Preparer Tax Identification numbers, and Corporate Affiliation files are all underway utilizing the ETL tool as an opportunity to complete projects more quickly and at a lower cost than older methods of coding. **On average, the coding effort has been reduced by 80 percent.**

The Vital Records business unit in the Michigan Department of Community Health had been using an old version of the software running on a desktop to perform **data matching on birth and death records to ensure quality analysis and forecasting**. They now save $15,000 per year in annual maintenance by switching to the enterprise version and have much faster processing and capacity available.

In addition, this project has incorporated test/development and disaster recovery environments as part of the architectural design allowing departments to **better protect their production environment** and more fully test functionality before deploying. Many smaller projects do not have sufficient resources to build out such a robust test and disaster recovery environment. This forced the development teams to handle production upgrades through complex workarounds on dissimilar platforms or to test changes on weekends using the production area. It is anticipated that the new test environments will **improve quality of the production ready code and significantly reduce the number of outages and job failures**.

Upgrades and patches to the environments are now handled centrally, assuring uniform application of changes and coordination of testing among all departments utilizing the services. This will ensure that development and viewing licenses will be synchronized within the application, avoiding issues that can occur when releases of the application software are incompatible with one another.

Additional benefits of this project include savings in terms of processing time (65% on average for the jobs converted from mainframe processes) and in development time (a

savings of approximately 80% on development time for the first several projects selected for the new toolset based on estimates). As stated above, a single Treasury project saved 1200 hours and over $200,000 from original projected estimates.

Currently the Department of Community Health is utilizing the tool to quickly develop a process to transfer data from an Oracle database directly into the Teradata data warehouse in support of the Maternal Infant Health Care process. The tool is also being used to develop matching and load tables for a study of the homeless population in Michigan, bringing data in from a number of different sources and agencies for analysis.

Financial Return - **Data quality improvements, based on responses from Treasury are noted as improvements ranging from 10-25% over existing processes. These improvements will show themselves in improved matching of tax non-compliance activities and recovery efforts. This will result in 500-1,500 fewer incorrect tax audits and a potential additional $13-$33 Million in tax revenue.**

Investment - Project costs to-date (including hardware, software licenses, set up, consultants, and staff) $823,047.90. Projected costs (including ongoing internal hosting, server replacement [4 year cycle], license maintenance, and on-staff administrator) are $280,000 per year.

Initial Investment		**Ongoing Costs**	
Licenses and upgrades	$390,132	Licenses	$ 60,000
Hardware	$208,271	Hardware Replacement	$ 60,000
Annual Hosting Costs	$ 8,263	Annual Hosting Costs	$ 10,000
Labor Incl. Vendor Asst.	$203,881	Tool Administration	$150,000*
Training	$ 12,500		
TOTAL	$823,047	TOTAL	$280,000

*Tool administration is fully loaded with benefits and represents 1 full time DBA.

Savings - Savings and recoveries realized to-date (including savings, amounts billed to departments, and internal grants) are $492,638. A five-year savings of $7,762,500 is estimated as follows.

Year	**2009**	**2010**	**2011**	**2012**	**2013**
Projected savings*	$582,000	$1,122,000	$1,523,500	$1,635,000	$1,770,000

*Savings are based upon 75% development time savings as well as conservative projections of usage. Also included is $15,000 in eliminated maintenance contracts.